Available online at www.sciencedirect.com

ScienceDirect

International Journal of Approximate Reasoning
44 (2007) 322–338

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Decision making under incomplete data using the imprecise Dirichlet model

L.V. Utkin [a,*], Th. Augustin [b]

[a] *Department of Computer Science, Forest Technical Academy, Institutski per. 5, 194021 St. Petersburg, Russian Federation*

[b] *Department of Statistics, University of Munich, Germany*

Received 15 December 2005; received in revised form 30 June 2006; accepted 31 July 2006
Available online 25 September 2006

Abstract

The paper presents an efficient solution to decision problems where direct partial information on the distribution of the states of nature is available, either by observations of previous repetitions of the decision problem or by direct expert judgements.

To process this information we use a recent generalization of Walley's imprecise Dirichlet model, allowing us also to handle incomplete observations or imprecise judgements, including missing data. We derive efficient algorithms and discuss properties of the optimal solutions with respect to several criteria, including Gamma-maximinity and *E*-admissibility. In the case of precise data and pure actions the former surprisingly leads us to a frequency-based variant of the Hodges–Lehmann criterion, which was developed in classical decision theory as a compromise between Bayesian and minimax procedures.

Keywords: Belief functions; Coarse data; Decision making; *E*-admissibility; Imprecise Dirichlet model (IDM); Imprecise probabilities; Incomplete data; Interval probability; Interval statistical models; Missing or set-valued statistical data

* Corresponding author.
E-mail addresses: lvu@utkin.usr.etu.spb.ru (L.V. Utkin), thomas@stat.uni-muenchen.de (Th. Augustin).

0888-613X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2006.07.016

1. Introduction

When applying the theoretical framework of decision theory to substantial science problems, decision makers typically have only limited information about probability distributions involved in the problem, and so the analysis is associated with large uncertainty. As a result, decision makers are frequently confronted with the problem that the very demanding and strong conditions of the classical probability calculus, and the decision models based on it, are not satisfied. Then it is indispensable to ask how to take into account the limitation of information and what conclusions can be drawn on the basis of such limited information.

Various tools for sophisticated uncertainty representation generalizing the common ('classical') concept of probability can be found in the literature, including Dempster–Shafer structures [18,42], interval-valued probabilities [51], imprecise probabilities [33,49], see also [6,12,13]. The corresponding decision making models have been developed in accordance with the different types of the uncertainty representation (e.g. [1,3,20,36,23, 37,43,47,53]). In contrast to standard decision theory, these models allow to handle *partial* information about the stochastic behavior of the states of nature.

Here we explicitly take into account the construction of the information and consider decision problems where direct data on the states are available. The data are of multinomial structure, consisting of independent categorical observations. As usual, real values can be associated with the observations as long as the ordering in these values is not understood as providing additional information. In addition, the model also seems to be suitable for processing expert judgements, as long as they are based on independent sources of information.

A particularly attractive feature of our method is that it will prove to be able to incorporate even set-valued observations, i.e., to handle situations where the corresponding category cannot be observed exactly and is only known to belong to a certain subset of the sample space. This is an important issue in many applications, but up to now there is no unique terminology. Depending on the context, different terms are common, like 'coarse data', or 'incomplete data', to denote such data sets as a whole, and 'imperfect measurement' or 'interval-valued observations', to denote the single set-valued observations.

To process complete multinomial data a Bayesian would recommend to use the Dirichlet model (for ease of distinction called *precise Dirichlet model* (PDM) in the sequel). The PDM has been widely adapted to many applications due to interesting statistical properties, in particular, due to the important fact that the Dirichlet density functions constitute a conjugate family of density functions with respect to multinomial likelihoods. A very promising generalization of the PDM, taking into account lack of prior information, is Walley's *imprecise Dirichlet model* (IDM), (cf. [50]; for a recent survey of applications see [7]).

This paper applies the IDM to decision making and derives simple algorithms for computing optimal randomized and pure actions. The method developed solves two practically important problems that cannot be addressed by any of the classical approaches to decision theory in a satisfying manner: first of all, relying on the IDM enables us to take into account explicitly that the number of judgements or measurements may be rather small, i.e. much small for being able to apply asymptotic arguments, based on the consistent estimation of the distribution of the states of nature. Secondly, we allow information about states of nature to be represented by imprecise, i.e., for instance, interval-valued observations or measurements. It turns out that this general case can be handled by considering a set of IDMs.

The paper is organized as follows: In Section 2 we formulate the problem under consideration more precisely. After having recalled some basic aspects of the imprecise Dirichlet model in Section 3, we apply it in Section 4 to the decision problem. In Section 5 we derive algorithms to determine the optimal randomized and unrandomized actions under a pessimistic criterion relying on strict ambiguity aversion. Section 6 extends consideration to imprecise observations and judgements. Close relations to Dempster–Shafer decision making will be illuminated, and a numerical example will be analyzed. Section 7 glances at more complex decision criteria and Section 8 concludes with some final remarks.

2. Statement of the decision problem

Consider the basic model of decision theory: one has to choose an *action* from a non-empty, finite set $\mathbb{A} = \{a_1, \dots, a_r, \dots, a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state of nature* $\omega \in \Omega = \{\omega_1, \dots, \omega_j, \dots, \omega_m\}$. The corresponding outcome is evaluated by the *utility function* $u : (\mathbb{A} \times \Omega) \to \mathbb{R}, (a, \omega) \mapsto u(a, \omega)$ and by the associated random variable $\mathbf{u}(a)$ on $(\Omega, \mathscr{P}o(\Omega))$ taking the values $u(a, \omega)$.[1] Often it makes sense to study *randomized actions*, which can be understood as a probability measure $\lambda = (\lambda_1, \dots, \lambda_r, \dots, \lambda_n)$ on $(\mathbb{A}, \mathscr{P}o(\mathbb{A}))$. Then $u(\cdot)$ and $\mathbf{u}(\cdot)$ are extended to randomized actions by defining $u(\lambda, \omega) := \sum_{r=1}^{n} u(a_r, \omega)\lambda_r$. Let Λ denote the set of all randomized actions and identify every pure action $a_r \in \mathbb{A}$ with the corresponding randomized action where $\lambda(a_r) = 1$ and $\lambda(a_s) = 0$, for all $s \neq r$.

This model contains the essentials of every (formalized) decision situation under uncertainty and is applied in a huge variety of disciplines. If the states of nature are produced by a perfect random mechanism (e.g. an ideal lottery), and the corresponding probability mass function $\pi(\cdot)$ on the sample space Ω is completely known, then the Bernoulli principle is almost unanimously favored.

Then one chooses the randomized action maximizing the expected utility

$$\mathbb{E}_\pi \mathbf{u}(a) := \sum_{j=1}^{m} u(a, \omega_j) \cdot \pi(\omega_j) \quad \text{and} \quad \mathbb{E}_\pi \mathbf{u}(\lambda) := \sum_{j=1}^{m} u(\lambda, \omega_j) \cdot \pi(\omega_j) \tag{1}$$

among all $a \in \mathbb{A}$ and all $\lambda \in \Lambda$, respectively. For simplicity, in the sequel the obvious constraints $\lambda_j \geqslant 0$ will be omitted in most places, and we often use the following abbreviated notation:

$$\pi_j := \pi(\omega_j), \quad u_{rj} := (a_r, \omega_j), \quad \mathbf{u}_r := \mathbf{u}(a_r), \quad \underline{\mathbf{u}}_r := \min_{j=1,\dots,m} u_{rj},$$
$$\overline{\mathbf{u}}_r := \max_{j=1,\dots,m} u_{rj}, \quad \mathbf{1} := (1, \dots, 1)^{\mathrm{T}}.$$

Here we aim at developing a powerful method to solve such decision problems when $\pi(\cdot)$ is not known, but – potentially imprecise – data from previous repetitions or expert judgements are available. If one had infinitely many – precise – observations, one would be able to apply the Bernoulli principle based on a consistent plug-in estimator of $\pi(\cdot)$. In

[1] Alternatively a loss function $l(a, \omega)$ is assigned, which can be embedded into the framework proposed by setting $u(a, \omega) = -l(a, \omega)$. $\mathscr{P}o$ denotes the power set.

applications, most often the sample size is too small to ignore the uncertainty in the estimation of π, and alternative methods are needed taking into account the lack of complete information explicitly. To achieve this, we base our proposal on Walley's imprecise Dirichlet model [50], and a recent generalization of it [46].

3. Walley's imprecise Dirichlet model (IDM)

The observations ω_j are assumed to be categorical, unordered and exchangeable. Then the corresponding multivariate random quantity n_j counting the number of occurrences of the jth category follows a multinomial distribution with parameter vector π. In a Bayesian setting the corresponding conjugated prior is the (precise) Dirichlet $(s,\mathbf{t})$ prior distribution (e.g. [17]), where $\mathbf{t}=(t_1,\ldots,t_m)$ is a certain element of the interior of the unit simplex denoted by $S(1,m)$. The parameter $t_i\in(0,1)$, $i=1,\ldots,m$ is the mean of π_i under the Dirichlet prior; the hyperparameter $s>0$ determines the influence of the prior distribution on posterior probabilities.

An important argument against the use of the PDM is that – at least without a huge amount of observations – inferences depend to a considerable extent on the value of $\mathbf{t}$ to be fixed in advance, typically without having sufficient information to guide the choice. Moreover, there is the desire for a model where the predictive probabilities used in decision making directly reflect the sample size, i.e. the amount of statistical information available.

Both problems are simultaneously addressed by the imprecise Dirichlet model as defined by Walley [50]. It avoids unjustifiable prior choices of $\mathbf{t}$ by relying on the set of *all* Dirichlet $(s,\mathbf{t})$ distributions such that $\mathbf{t}\in S(1,m)$, and the inferences depend – via the width of the intervals for the predictive probability obtained – on the sample size. In the IDM, there is a *hyperparameter* s determining how quickly upper and lower probabilities of events converge as statistical data accumulate. s can be interpreted as either the number of observations needed to reduce the imprecision (i.e. the difference between upper and lower probabilities) to half its initial value, or alternatively as the virtual number of yet unseen observations. Consequently, smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. At the same time, according to Walley, the value of s must not depend on m or the number of observations. A detailed discussion concerning the parameter s and the IDM can be found in particular in [7,50].

To derive the predictive lower and upper probabilities assigned by the IDM, let A be any non-trivial subset of the sample space $\Omega=\{\omega_1,\ldots,\omega_m\}$, i.e., $A\neq\emptyset$ and $A\neq\Omega$, and let $n(A)$ denote the observed number of occurrences of A in N trials, $n(A)=\sum_{\omega_j\in A}n_j$ where $n_j:=n(\{\omega_j\})$. Then the predictive probability $P(A|\mathbf{n},\mathbf{t},s)$ under a certain Dirichlet posterior distribution is

$$P(A|\mathbf{n},\mathbf{t},s)=\frac{n(A)+st(A)}{N+s},$$

where $t(A):=\sum_{\omega_j\in A}t_j$. This assignment is completed by taking $P(A|\mathbf{n},\mathbf{t},s):=0$ if A is empty, and $P(A|\mathbf{n},\mathbf{t},s):=1$ if $A=\Omega$. By maximizing and minimizing $P(A|\mathbf{n},\mathbf{t},s)$ over $\mathbf{t}\in S(1,m)$, Walley [50] obtains the posterior lower and upper predictive probabilities of A as

$$\underline{P}(A|\mathbf{n},s)=\frac{n(A)}{N+s},\quad \overline{P}(A|\mathbf{n},s)=\frac{n(A)+s}{N+s}.$$

An illuminating non-Bayesian view of this model as ϵ-contaminated relative frequencies is provided by [41].

4. Decision making by using the imprecise Dirichlet model

As a preparation, let us briefly consider an approach for decision making under the condition that π satisfies to the PDM and there are perfect observations of states of nature, namely the numbers $(n_1,\ldots,n_m)$ of occurrences of $\omega_1,\ldots,\omega_m$ in the N trials. By relying on Bayesian methodology, the expected utility of an action λ is calculated as follows:

$$\begin{aligned}\mathbb{E}\mathbf{u}(\lambda)&=\int_{S(1,m)}\sum_{i=1}^{m}(u(\lambda,\omega_i)\cdot\pi_i)p(\pi)\,\mathrm{d}\pi=\sum_{i=1}^{m}u(\lambda,\omega_i)\cdot\int_{S(1,m)}\pi_i p(\pi)\,\mathrm{d}\pi\\&=\sum_{i=1}^{m}u(\lambda,\omega_i)\cdot\mathbb{E}_p\pi_i,\end{aligned}$$

where $\mathbb{E}_p\pi_i=\frac{n_i+st_i}{N+s}$, finally resulting in

$$\mathbb{E}\mathbf{u}(\lambda)=\sum_{i=1}^{m}u(\lambda,\omega_i)\frac{n_i+st_i}{N+s}. \tag{2}$$

Passing over to the IDM leads to lower and upper expected utilities arising from the following optimization problems:

$$\underline{\mathbb{E}}\mathbf{u}(\lambda)=\inf_{\mathbf{t}\in S(1,m)}\mathbb{E}\mathbf{u}(\lambda),\quad \overline{\mathbb{E}}\mathbf{u}(\lambda)=\sup_{\mathbf{t}\in S(1,m)}\mathbb{E}\mathbf{u}(\lambda). \tag{3}$$

In literature several criteria have been suggested to compare the interval-valued expected utility

$$[\underline{\mathbb{E}}\mathbf{u}(\lambda),\overline{\mathbb{E}}\mathbf{u}(\lambda)] \tag{4}$$

of actions λ (see, in particular, the recent survey by [45] as well as [48], who both give further references).[2] Two types of criteria may be distinguished with respect to their ordering properties. The first branch, like the criterion of maximality (as proposed by [49]) or the concept of E-admissibility (advocated by [34,38]), renounces the completeness of the ordering and generalizes the concept of admissibility by distinguishing a set of actions as being undominated, i.e. not inferior.

On the other hand, often a complete ordering of the actions is desired, and to achieve this the interval-valued expected utility eventually has to be transformed to the real line. The most conservative choice is to be strictly ambiguity averse, concentrating on the lower interval limit only. (Sections 5 and 6 will rely on this criterion, while Section 7 briefly will consider alternative criteria.)

[2] For rigorous axiomatic justifications of generalized expected utility in the sense of (4) as well as different criteria derived from it see among others [21,22] and the work cited there. Relying on these results it is possible to extend Neumann–Morgenstern and Anscombe–Aumann theory to the situation of complex uncertainty with partial prior information.

5. Decision making under strict ambiguity aversion

Under strict ambiguity aversion one tries to be safe in the worst situation, i.e. one evaluates every action by its minimal expected utility among all probabilities in accordance with $\underline{P}(\cdot)$ and $\overline{P}(\cdot)$. Consequently, an action λ^* is optimal iff for all λ the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geqslant \underline{\mathbb{E}}\mathbf{u}(\lambda)$ is satisfied.[3] An algorithm to calculate λ^* is described in

Proposition 1. *The optimal randomized action λ^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geqslant \underline{\mathbb{E}}\mathbf{u}(\lambda)$ for all λ is determined by solving the following linear programming problem*:

$$G \to \max_{\lambda,G}{}^{4} \tag{5}$$

subject to $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$, *and for* $j = 1, \ldots, m$,

$$G \leqslant \frac{1}{N+s} \sum_{r=1}^{n} \lambda_r \left(s \cdot \underline{\mathbf{u}}_r + \sum_{i=1}^{m} u_{ri} \cdot n_i \right). \tag{6}$$

Proof. It follows from (2) that λ^* is found by considering

$$\inf_{\mathbf{t}\in S(1,m)} \sum_{i=1}^{m} \sum_{r=1}^{n} u_{ri}\lambda_r \cdot \frac{n_i + st_i}{N+s} \to \max_{\lambda} \tag{7}$$

subject to $\lambda \cdot \mathbf{1} = 1$. For solving this problem, let us adapt [1,3], who suggested to introduce a new variable $G = \inf_{\mathbf{t}\in S(1,m)} \mathbb{E}\mathbf{u}(\lambda)$. Then problem (7) is equivalent to a problem with objective function (5) and with the infinite number of constraints

$$G \leqslant \sum_{i=1}^{m} \sum_{r=1}^{n} u_{ri}\lambda_r \cdot \frac{n_i + st_i}{N+s}, \quad \mathbf{t} \in S(1,m), \tag{8}$$

and $G \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$. Following [1,3] further, note that the constraints are already satisfied, if they are satisfied for all extreme points of the convex sets of distributions defined by $\underline{P}(\cdot|\mathbf{n},s)$ and $\overline{P}(\cdot|\mathbf{n},s)$, which are simply obtained by considering the extreme points of $S(1,m)$. The latter have the form $(1,0,\ldots,0), (0,1,\ldots,0), \ldots, (0,0,\ldots,1)$.[5] Therefore, (8) is reduced to the set of m linear constraints described in (6), as was to be proven.[6] $\square$

Restricting λ_r to be either 0 or 1 optimal unrandomized actions are obtained via

[3] This criterion has been proposed under different names. It corresponds to the Gamma-Minimax criterion (as considered, e.g., in [5, Section 4.7.6]), to the Maxmin expected utility model [24], to the MaxEMin criterion investigated by [31] (cf. also [32] and the references therein) and the notion of maximinity in [49]. In the case of two-monotone capacities it is equivalent to maximizing Choquet expected utility (as studied, e.g., in [9]).

[4] This abbreviation, which is used throughout the paper, stands for task to maximize the objective function G with respect to the variables λ and G, subject to the constraints that are following.

[5] To be exact, the extreme points are of the form $(1-(m-1)\cdot\varepsilon, \varepsilon, \ldots, \varepsilon)$, where ε becomes arbitrarily small.

[6] An alternative proof can rely on total monotonicity of the lower interval limit derived from the IDM and the equivalence shown at the beginning of Section 6.2 and the formula from footnote 8, again circumventing minima by auxiliary variables.

Corollary 1. *The optimal unrandomized action (pure action) a_k satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}_k \geqslant \underline{\mathbb{E}}\mathbf{u}_r$, for all $r = 1, \dots, n$, is determined by solving the following problem:*

$$\frac{1}{N+s}\left(s \cdot \underline{\mathbf{u}}_r + \sum_{i=1}^{m} u_{ri} n_i\right) \to \max_r, \quad r \in \{1, \dots, n\}. \tag{9}$$

It can be seen from (9) that the objective function is nothing else but a mixture of two criteria: the criterion of maximum expected utility, with probabilities $\pi(\{\omega_i\})$ replaced by the corresponding relative frequency n_i/N, and Wald's criterion. The weights $N/(N+s)$ and $s/(N+s)$, respectively, are directly connected to the uncertainty involved, depending on the hyperparameter s and the sample size N. Consequently, when $N = 0$, i.e., before any observation on the states of nature is available, Wald's criterion is used. On the other hand, when the sample size tends to infinity, i.e. when there are enough data to estimate $\pi(\cdot)$ with very low variance, expected utility based on the estimate of $\pi(\cdot)$ is used. In this sense, (9) is a frequency-based type of Hodges–Lehmann criterion (cp. [28]), which has been proposed in classical decision theory as a compromise between the Bayesian and the minimax approach. As a welcome by-product, Eq. (9) also provides a behavioral interpretation of the hyperparameter s: In principle, one could develop canonical examples to determine a decision maker's value of s in an experimental manner.

A closer investigation of (9) shows, however, also a possibly unwanted effect: When the utility function is such that $\underline{\mathbf{u}}_r$ is the same for all actions a_r, then the first summand in (9) does not matter and $\frac{\sum u_{ri} \cdot n_i}{N+s}$ is maximized, which distinguishes the same action as optimal as the naive frequentist approach where in (7) the probability $\pi(\cdot)$ is replaced by the vector of observed relative frequencies. Following up on this matter, note that, in principle, such a situation can always be constructed, by adding a "bad" state of nature ω_0 (like "crash of the economic system" in the example in Section 6.4) that has constant utility u_0 for all actions such that $u_0 < \min_{r,i} u_{ri}$. This means that – just as in minimax theory – special attention has to be paid to careful selection of the states of nature that are taken into consideration. To turn it into other – even more trenchant – words: The representation invariance principle (RIP), which is understood as being crucial for the IDM, cannot be extended to decision making in a straightforward manner. As the RIP says, the lower and upper posterior and predictive probabilities of some event A do not depend on the sample space in which A and the previous observations are presented, but, as just argued, the optimal action may do so.

6. Decision making under incomplete data

Now consideration is extended to the practically quite important case of imprecise observations, where observations may be too vague to be associated with a certain singleton $\{\omega_j\}$. Instead it is only known that the realized state of nature lies in some subset $A_i \subseteq \Omega$, see, for instance [27] for examples in the context of biometrics. Heitjan and Rubin, who have coined the term coarsened for such data, derive in [26] – rather severe – conditions under which the coarsening may be handled in an easy way. Blumenthal (cf. [8]) discusses a multinomial model under the additional assumption that the probability distribution of the coarsening process is known.

However, quite often Heitjan and Rubin's so-called coarsening at random assumption is violated and the naive, straightforward analysis would be heavily biased. The same applies, as the simulations in [35] show, when the typically unknown distribution in Blumenthal's model is misspecified. Therefore a thorough analysis without relying on unjustified assumptions is highly desirable (see, in particular [55] for a general framework). Several authors, among them [44], have understood Dempster–Shafer belief functions/random sets as an appropriate tool to model such situations: they use so-to-say empirical belief functions (see below) based on relative frequencies of the observed subsets A_i for the analysis. Although then incompleteness in the observations is taken into account, still a severe bias may occur, because this way to proceed neglects – by implicitly equating relative observed frequencies and probabilities – finite sample variation, which may have, as argued above and later, a strong distorting effect, too.

6.1. Extended empirical belief functions

In order to handle both sources of potential bias – imprecision in the observations as well as the limited sample size – we rely on a model recently developed by [46], which, in essence, considers all multinomial models compatible with the data and will lead to a powerful extension of empirical belief functions.

To be a bit more precise (for a detailed account the reader is referred to [46]): Data consist now of c_i observations of the non-empty subset $A_i \subseteq \Omega$, $i = 1, \ldots, M$, such that $\sum_{i=1}^{M} c_i = N$. Furthermore, it is helpful to introduce sets J_i denoting the set of indices of states of nature belonging to A_i, i.e. $A_i = \{\omega_j : j \in J_i\}$. Evidently the standard case is included: there all available subsets consist of singletons, i.e., $M = m$, $A_j = \{\omega_j\}$ and $J_j = \{j\}, j = 1, \ldots, m$. The case $A_{i_0} = \Omega$ for some i_0 represents in a straightforward manner cautious treatment of missing data.

Translating the situation into an urn model, we have m urns $\omega_1, \ldots, \omega_m$ (corresponding to the states of nature), where ω_j consists of balls with number j. Relying on the notation just introduced, we randomly choose subsets A_i of urns and take randomly c_i balls from the urns numbered by elements of J_i. Given M, there exist different possible combinations $k = 1, \ldots, K$ of numbers of balls taken from the urns. Denote the kth possible vector of balls by $\mathbf{n}^{(k)} = (n_1^{(k)}, \ldots, n_m^{(k)})$, where n_l^k, $l = 1, \ldots, m$, is the number of balls with number l drawn in the M choices, and let $\mathbf{c} = (c_1, \ldots, c_n)$. Assuming that the subsets A_i are independently chosen from the set of all subsets of Ω and that the probability of selecting a ball from the jth urn is π_j, every combination of balls produces one standard multinomial model. A number of possible combinations of balls produce the same number of standard multinomial models. Moreover, we cannot prefer one model over another.

Since we have a set of vectors $\mathbf{n}^{(k)}$, then even if we know precisely the predictive probabilities $P(A|\mathbf{n}^{(k)})$ for every event A in Ω and every possible vector $\mathbf{n}^{(k)}$, still we can only compute lower and upper probabilities for events A:

$$\underline{P}(A|\mathbf{c}) = \min_k P(A|\mathbf{n}^{(k)}), \quad \overline{P}(A|\mathbf{c}) = \max_k P(A|\mathbf{n}^{(k)}).$$

As the vectors $\mathbf{n}^{(k)}$ depend on $\mathbf{c}$, the resulting lower and upper probabilities (after minimizing and maximizing $P(A|\mathbf{n}^{(k)})$) depend on $\mathbf{c}$, and so it makes indeed sense to denote them by $\underline{P}(A|\mathbf{c})$ and $\overline{P}(A|\mathbf{c})$.

Following the same argumentation as above, using the IDM allows us – in contrast to the PDM – to take into account lack of prior information and the possible fact that the number of observations may be rather small. Then the lower and upper probabilities

$\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ of an arbitrary event A, corresponding to a index set $J \subseteq \{1,\ldots,m\}$, are computed as follows:

$$\underline{P}(A|\mathbf{c},s) = \frac{\min_k n^{(k)}(A) + s \cdot \inf_{\mathbf{t}\in S(1,m)} t(A)}{N+s},$$
$$\overline{P}(A|\mathbf{c},s) = \frac{\max_k n^{(k)}(A) + s \cdot \sup_{\mathbf{t}\in S(1,m)} t(A)}{N+s},$$

where $t(A) := \sum_{j\in J} t_j$ and $n^{(k)}(A) := \sum_{j\in J} n_j^{(k)}$.

It can be shown that the resulting lower and upper probabilities of A can be obtained from the observations $A_1,\ldots,A_M$ and $c_1,\ldots,c_M$ as follows:

$$\underline{P}(A|\mathbf{c},s) = \frac{\sum_{i:A_i\subseteq A} c_i}{N+s}, \quad \overline{P}(A|\mathbf{c},s) = \frac{\sum_{i:A_i\cap A\neq\emptyset} c_i + s}{N+s}. \tag{10}$$

For a closer investigation of these results, it helps to consider them within the framework of Dempster–Shafer theory (e.g. [18,42]): With [35] we call a basic probability assignment $m : \mathscr{P}o(\Omega) \to [0,1]$, with $m(\emptyset) = 1$, $\sum_{A\in\mathscr{P}o(\Omega)} m(A) = 1$ and the corresponding belief $\mathrm{Bel}(A)$ and plausibility $\mathrm{Pl}(A)$ functions with $\mathrm{Bel}(A) = \sum_{B\subseteq A} m(B)$, $\mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^c)$ *empirical*, when

$$m(A_i) = \frac{c_i}{N}$$

based on a vector $c_1,\ldots,c_M$ of $\sum_{i=1}^M c_i = N$ observations of $A_i \subseteq \Omega$, $i = 1,\ldots,M$, and $m(B) = 0$ for the remaining events $B \in \Omega$. It is easy to see that the lower and upper probabilities (10) relate to these belief and plausibility functions in the following way:[7]

$$\underline{P}(A|\mathbf{c},s) = \frac{N \cdot \mathrm{Bel}(A)}{N+s}, \quad \overline{P}(A|\mathbf{c},s) = \frac{N \cdot \mathrm{Pl}(A) + s}{N+s}.$$

Moreover, $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ are belief and plausibility functions again, namely with the basic probability assignment $m^*(A_i) = c_i/(N+s)$ for every A_i, $m(B) = 0$ for the remaining sets $B \subsetneq \Omega$ and $m^*(A_\Omega) = s/(N+s)$, i.e., $\underline{P}(A|\mathbf{c},s)$ and $\overline{P}(A|\mathbf{c},s)$ can be obtained as empirical belief and plausibility functions by assuming that there are s additional observations Ω. With $m(A_i) = c_i/N$ representing the standard empirical assignment, we have $m^*(A_i) = m(A_i) \cdot N/(N+s)$, and, for all $A \neq \Omega$,

$$\underline{P}(A|\mathbf{c},s) = \sum_{i:A_i\subseteq A} m^*(A_i),$$
$$\overline{P}(A|\mathbf{c},s) = \sum_{i:A_i\cap A\neq\emptyset} m^*(A_i),$$

which we call *extended empirical* belief and plausibility functions.

6.2. Decision making with extended empirical belief functions

Again two ways to use such information in decision making suggest themselves: The first one uses the extended empirical belief functions from the previous subsection, and

[7] A detailed study of extended empirical belief functions derived from the imprecise Dirichlet model is presented in [46].

looks for every randomized action λ at the interval-valued expected utility $[\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda), \overline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda)]$ with

$$\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) := \inf_{p\in\mathscr{M}} \mathbb{E}_p\mathbf{u}(\lambda) \quad \text{and} \quad \overline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) := \sup_{p\in\mathscr{M}} \mathbb{E}_p\mathbf{u}(\lambda),$$

where $\mathscr{M}$ is the corresponding structure, i.e. the set of classical probabilities $\pi(\cdot)$ dominating $\underline{P}(\cdot)$, i.e.

$$\mathscr{M} = \{\pi(\cdot)|\underline{P}(A|\mathbf{c}, s) \leqslant \pi(A)\}. \tag{11}$$

In this context it is very convenient to note that extended empirical belief functions are still belief functions, and so we can use an approach based on the Choquet integral [10,43], which directly relies on basic probability assignment $m^*(\cdot)$.[8] In the situation under consideration described by the basic probability assignment $m^*(\cdot)$, the lower expected utility can be rewritten as follows:

$$\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) = \frac{s}{N+s} \cdot \min_{\omega_i\in\Omega} u(\lambda,\omega_i) + \sum_{k=1}^{M} \frac{c_k}{N+s} \cdot \min_{\omega_i\in A_k} u(\lambda,\omega_i). \tag{12}$$

The second way directly understands the situation as a collection of imprecise Dirichlet models, where each of them is associated with an expected utility according to (3). Taking the lower and upper envelope yields $[\underline{\mathbb{E}}\mathbf{u}(\lambda), \overline{\mathbb{E}}\mathbf{u}(\lambda)]$ with

$$\underline{\mathbb{E}}\mathbf{u}(\lambda) = \min_k \inf_{\mathbf{t}\in S(1,m)} \mathbb{E}^{(k)}\mathbf{u}(\lambda) = \min_k \inf_{\mathbf{t}\in S(1,m)} \sum_{i=1}^{m} u(\lambda,\omega_i) \cdot \frac{n_i^{(k)} + st_i}{N+s}. \tag{13}$$

In general,

$$\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) \leqslant \underline{\mathbb{E}}\mathbf{u}(\lambda) \leqslant \overline{\mathbb{E}}\mathbf{u}(\lambda) \leqslant \overline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda), \tag{14}$$

but both ways need not coincide (use e.g. the example from [49, p. 82ff.]). Here, however, we obtain

Proposition 2. *Comparing (12) and (13) yields*

$$\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) = \underline{\mathbb{E}}\mathbf{u}(\lambda) \quad \textit{and} \quad \overline{\mathbb{E}}\mathbf{u}(\lambda) = \overline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda). \tag{15}$$

Proof. According to (14), it is sufficient to show $\underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda) \geqslant \underline{\mathbb{E}}\mathbf{u}(\lambda)$. This can be achieved by constructing a constellation where $\mathbb{E}^{(k)}\mathbf{u}(\lambda) = \underline{\mathbb{E}}_{\mathscr{M}}\mathbf{u}(\lambda)$. For this purpose consider that vector $\mathbf{n}$ where for every A_k the corresponding c_k is completely assigned to that state, where $u(\lambda,\omega_i)$ is minimal. The corresponding expectation coincides with (12). □

Efficient handling of the resulting decision problem is summarized in

[8] More precisely, two-monotonicity of the lower probability $\underline{P}(\cdot)$ and the upper probability $\overline{P}(\cdot)$ would be sufficient to represent expectations in the spirit of [11] via the Moebius inverse of $\underline{P}(\cdot)$ (cf. [10, Corollary 4]), yielding for any random variable X on Ω

$$\underline{\mathbb{E}}X = \sum_{A\subseteq\Omega} m(A) \cdot \min_{\omega\in A} X(\omega).$$

.

Proposition 3. *If the probabilities of m states of nature are described by the imprecise Dirichlet model with the hyperparameter s and information about the states is represented in the form of c_i observations of subsets $A_i = \{\omega_j : j \in J_i\}$, $i = 1,\dots,M$, such that $\sum_{i=1}^{M} c_i = N$, then the optimal randomized action λ^* satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}(\lambda^*) \geqslant \underline{\mathbb{E}}\mathbf{u}(\lambda)$ for all λ is determined by solving the following linear programming problem:*[9]

$$\frac{1}{N+s}\left(s \cdot V_0 + \sum_{k=1}^{M} c_k \cdot V_k\right) \to \max_{\lambda}$$

subject to $V_0, V_i \in \mathbb{R}$, $\lambda \cdot \mathbf{1} = 1$,

$$V_i \leqslant \sum_{r=1}^{n} \lambda_r u_{rj}, \quad i = 1,\dots,M, \ j \in J_i, \qquad V_0 \leqslant \sum_{r=1}^{n} \lambda_r u_{rj}, \quad j = 1,\dots,m.$$

Proof. We introduce new variables $V_i = \min_{j \in J_i} u(\lambda, \omega_j)$, $i = 1,\dots,M$, and $V_0 = \min_{j=1,\dots,m} u(\lambda, \omega_j)$, and substitute them into the objective function. Constraints of the optimization problem are derived from the definition of V_i, $i = 0,\dots,M$. □

The extension of Corollary 1 to the situation under consideration is provided by

Corollary 2. *The optimal unrandomized action (pure action) a_r satisfying the inequality $\underline{\mathbb{E}}\mathbf{u}_r \geqslant \underline{\mathbb{E}}\mathbf{u}_k$ for all $k = 1,\dots,n$ is determined by solving:*

$$\frac{1}{N+s}\left(s \cdot \underline{\mathbf{u}}_r + \sum_{k=1}^{M} c_k \cdot \min_{\omega_i \in A_k} u_{ri}\right) \to \max_{r}, \quad r \in \{1,\dots,,n\}. \tag{16}$$

Proof. If $\lambda_r \in \{0,1\}$, then $V_k = \min_{\omega_i \in A_k} u_{ri}$ and $V_0 = \underline{\mathbf{u}}_r$ due to conditions $s \geqslant 0$ and $c_k \geqslant 0$. □

Before we illustrate our approach by a short example in Section 6.4, some additional remarks may be appropriate. Firstly, we mention an equivalent alternative to proceed. After that, in Section 6.3, the advantages of the extended modelling will be demonstrated in some extreme cases.

Remark 1. Since the information about states of nature is represented by means of lower $\underline{P}(A|\mathbf{c},s)$ and upper $\overline{P}(A|\mathbf{c},s)$ probabilities of all events $A \in \Omega$, the decision problem can also be solved by means of the approach proposed by [1,3], where the extreme points needed can be directly derived from the corresponding basic probability assignment $m^*(\cdot)$.[10]

[9] The problem considered in Proposition 3 has $\sum_{i=1}^{M} |J_i| + m + n + 1$ linear constraints and $2m + 1$ optimization variables. (Here $|J_i|$ denotes the cardinality of the set J_i.) If all the subsets A_i consist of the single elements ω_i, resulting in $c_i = n_i$, and $M = m$, then we get back, after simple transformations, the special case studied in Section 5.

[10] As another variant, the resulting lower and upper probabilities can be interpreted as arising from a special type of a generalized basic probability assignment in the sense of [4] (see also [14]), constructed from the set of IDMs that are derived from the different possible observations.

6.3. Comparison with standard empirical belief functions

One of the main shortcomings of using standard empirical belief functions in decision making is that they assign zero probabilities to yet unobserved states of nature. If we had an infinite number of observations, then the fact of zero probabilities could be accepted. However, if we have a finite number of observations (and more often quite a small number), then this fact may indeed lead to controversial results, as can be seen in a most pronounced way by considering the following decision problem: $\mathbb{A} = \{a_1, a_2\}$, $\Omega = \{\omega_1, \omega_2\}$, the utility function is $u_{11} = -1000$, $u_{12} = 1$, $u_{21} = u_{22} = 0$. Suppose that there is only one judgment ($M = 1$) such that $A_1 = \{\omega_2\}$. Using standard belief functions, which means to rely on $s = 0$, we obtain $\underline{\mathbb{E}}\mathbf{u}_1 = 1$ and $\underline{\mathbb{E}}\mathbf{u}_2 = 0$. Hence the optimal action is a_1, i.e. under almost complete ignorance, where we naturally intend to be cautious, we make the optimistic decision because the model using standard empirical belief functions concludes that the probability of state ω_1 is 0, and so it acts as if this state could never be observed. However, if we take $s > 0$, say $s = 1$, then $\underline{\mathbb{E}}\mathbf{u}_1 = (-1000 + 1)/2 = -499.5$ and $\underline{\mathbb{E}}\mathbf{u}_2 = 0$. Hence the optimal action is a_2, showing that the IDM indeed provides a way to avoid such over-optimistic reasoning.

A related, again rather problematic, issue of standard empirical belief functions is that the assignments do not depend on the sample size and therefore – the potentially very high – finite sampling variation is neglected: Consider, for instance, two samples taken from the same sample space, one with $n = 2$, the other one with $n = 200{,}000$.[11] If in the first case $n_1 = 1 = n_2$, and in the second one $n_1 = 100{,}000 = n_2$, then the relative frequencies for the states 1 and 2 are 1/2 each, in both cases, and therefore also the standard belief functions derived from them are the same, not distinguishing these substantially different situations. In contrast, the extended belief function approach takes into account that the second information is build on a much stronger basis: the imprecision in the assignment arising from the second situation is much less than that from the first one.

The problem becomes in particular drastic, when we return to a situation with unobserved states of nature and consider again the numerical example given above. If we have l identical observations of the second state of nature, then

$$\underline{\mathbb{E}}\mathbf{u}_1 = -\frac{1000 \cdot s}{l+s} + \frac{l}{l+s}, \quad \underline{\mathbb{E}}\mathbf{u}_2 = 0.$$

When $s = 0$ then neither $\underline{\mathbb{E}}\mathbf{u}_1 = 1$ nor $\underline{\mathbb{E}}\mathbf{u}_2 = 0$ depend on l. That is, our decision a_2 is the same irrespective of having 1 single observation (almost complete ignorance) or 10,000 identical observations (sufficient statistical data). If we take $s = 1$, then a_2 is only superior when $l > 1000$.

6.4. Numerical example

Consider the following toy example of a simplified investment decision. The states of nature are the states of economy during one year: growth – ω_1, medium growth – ω_2,

[11] To focus on the essential point of the argument we further assume in this example that all observations are precise.

Table 1
Values of the utility function u_{rj}

Actions	States of nature			
	ω_1	ω_2	ω_3	ω_4
Bonds a_1	6	9	9	8
Stocks a_2	12	7	3	−2
Deposit a_3	7	7	7	7

no change – ω_3, recession – ω_4. The problem is to choose one of the three actions with the given rates of return as shown in the body of Table 1.

Suppose that three experts, relying on independent sources of information,[12] supply the following judgments concerning the states of economy: two experts ($c_1 = 2$) believe that the state of economy will be "growth" or "medium growth" ($A_1 = \{\omega_1, \omega_2\}$), one expert ($c_2 = 1$) supposes that the state of economy will be "no" or "medium growth" ($A_2 = \{\omega_2, \omega_3\}$). So, $M = 2$, $N = 3$.

Let us find the optimal randomized action maximizing the lower expected utility under the condition $s = 1$. By using Proposition 3, we arrive at the following problem:

$$\frac{1}{3+1}(1 \cdot V_0 + 2 \cdot V_1 + 1 \cdot V_2) \to \max_{\lambda, V_1, V_2, V_3}$$

subject to $\lambda \in \mathbb{R}^3_+$, $V_i \in \mathbb{R}$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and

$$\begin{aligned}
&V_1 \leqslant 6\lambda_1 + 12\lambda_2 + 7\lambda_3, \quad V_0 \leqslant 6\lambda_1 + 12\lambda_2 + 7\lambda_3,\\
&V_1 \leqslant 9\lambda_1 + 7\lambda_2 + 7\lambda_3, \quad V_0 \leqslant 9\lambda_1 + 7\lambda_2 + 7\lambda_3,\\
&V_2 \leqslant 9\lambda_1 + 7\lambda_2 + 7\lambda_3, \quad V_0 \leqslant 9\lambda_1 + 3\lambda_2 + 7\lambda_3,\\
&V_2 \leqslant 9\lambda_1 + 3\lambda_2 + 7\lambda_3, \quad V_0 \leqslant 8\lambda_1 - 2\lambda_2 + 7\lambda_3.
\end{aligned}$$

Hence $V_0 = 27/4$, $V_1 = 27/4$, $V_2 = 33/4$, $\lambda_1^* = 7/8$, $\lambda_2^* = 1/8$, $\lambda_3^* = 0$. The optimal lower expected utility is 7.125.

Let us now find the optimal unrandomized action: by using (16) we get $\underline{\mathbb{E}}\mathbf{u}_1 = 6.75$, $\underline{\mathbb{E}}\mathbf{u}_2 = 3.75$, $\underline{\mathbb{E}}\mathbf{u}_3 = 7$. This implies that the third action is optimal.

It could be noted that, by taking $s = 0$, the optimal randomized action would be $\lambda_1^* = 5/8$, $\lambda_2^* = 3/8$, $\lambda_3^* = 0$, with an optimal lower expected utility of 7.75. At a first glance one might be tempted to say that the decision based on $s = 0$ would be better than that based on $s > 0$ because the lower expected utility in the case $s = 0$ is larger than the lower expected utility based on $s = 1$. However, as discussed in the previous subsection, this decision is incautious; the model in which it is optimal neglects the fact that the number of judgments is very small ($N = 3$). The optimal unrandomized action taking $s = 0$ is not unique because $\underline{\mathbb{E}}\mathbf{u}_1 = \underline{\mathbb{E}}\mathbf{u}_3$ and $\underline{\mathbb{E}}\mathbf{u}_2 < \underline{\mathbb{E}}\mathbf{u}_1$ in this case.

7. Other optimality criteria

In this paper up to now only one particular – quite pessimistic – optimality criterion has been studied. As a considerable improvement a more complex criterion of decision making

[12] The assumption of independence is crucial for direct applications of the multinomial likelihood underlying the IDM. Relaxations of this assumption are currently under investigation.

based on a linear combination of lower and upper expectations with the so-called *caution* parameter $\eta \in [0,1]$ (cp., e.g. [19,39,51]) can be considered. The caution parameter reflects the degree of ambiguity aversion; the more ambiguity averse the decision maker is, the higher is the influence of the lower interval limit of the generalized expected utility. $\eta = 1$ corresponds to strict ambiguity aversion, $\eta = 0$ expresses maximal ambiguity seeking attitudes. Methods for the choice of η are considered in detail by [39; 51, Chapter 2.6]. Relying on this criterion, a pure action a_k is distinguished as optimal iff for all $r \in \{1,\dots,n\}$

$$\eta\underline{\mathbb{E}}\mathbf{u}_k + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_k \geqslant \eta\underline{\mathbb{E}}\mathbf{u}_r + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_r.$$

The lower expected utility $\underline{\mathbb{E}}\mathbf{u}_r$ is computed by means of (16). The upper expected utility $\overline{\mathbb{E}}\mathbf{u}_r$ can be found in the same way, leading to the expression

$$\begin{aligned}\eta\underline{\mathbb{E}}\mathbf{u}_r + (1-\eta)\overline{\mathbb{E}}\mathbf{u}_r = {} & \frac{s}{N+s}(\eta\underline{\mathbf{u}}_r + (1-\eta)\overline{\mathbf{u}}_r) \\ & + \frac{1}{N+s}\left(\sum_{i=1}^{M} c_i\left(\eta \min_{j\in J_i} u_{rj} + (1-\eta)\max_{j\in J_i} u_{rj}\right)\right).\end{aligned}$$

It can be seen that the algorithm for computing the optimal pure action taking into account both lower and upper expected utilities is similar to the approach proposed in Corollary 1. However, it should be noted that the randomized action cannot be found in the same simple way; efficient algorithms for solving this problem are currently investigated (see also, in the light of Remark 1, Section 4 of [48]).

Of course, also criteria not necessarily producing a linear ordering of the actions deserve attention (cf. Section 4). [30; 48, Section 5] propose, independently of each other, closely related methods to determine E-admissible actions, which can – again referring to Remark 1 – also be used here: A pure action a_i can be shown to be E-admissible, if and only if, with $\mathcal{M}$ from (11),

$$\Pi_i = \left\{\pi(\cdot) \in \mathcal{M} \,\middle|\, \sum_{j=1}^{m} u(a_i,\vartheta_j)\pi(\vartheta_j) \geqslant \sum_{j=1}^{m} u(a_l,\vartheta_j)\pi(\vartheta_j), \forall l = 1,\dots,n\right\} \neq \emptyset,$$

i.e., if and only if the linear programming problem

$$\begin{aligned}& z \to \max_{(\pi^T,z)^T} \\ & \sum_{j=1}^{m} u(a_i,\vartheta_j)\pi(\vartheta_j) \geqslant \sum_{j=1}^{m} u(a_l,\vartheta_j)\pi(\vartheta_j), \quad \forall l = 1,\dots,n, \\ & \sum_{j=1}^{m} \pi(\vartheta_j) = z, \quad z \leqslant 1, \qquad \pi(\vartheta_j) \geqslant 0, \quad j = 1,\dots,m, \\ & \underline{P}(A|\mathbf{c},s) \leqslant \sum_{\omega\in A} \pi(\omega) \leqslant \overline{P}(A|\mathbf{c},s) \quad \forall A \subseteq \Omega\end{aligned} \tag{17}$$

has an optimal solution with $z = 1$. Ref. [48] also extends this algorithm to determine optimal randomized actions, which are, if admissibility is assured (e.g. by $\underline{P}(\{\omega\}|\mathbf{c},s) > 0$, $\forall\omega \in \Omega$) also maximal.

8. Concluding remarks

A method for decision making under imprecise information using the IDM has been proposed in this paper. This approach can also be regarded as a particular extension of the procedure relying on empirical belief functions. The considered special cases and the numerical example have shown that the method is reasonable even in cases where the number of possible imperfect measurements or judgements is very small.

De Cooman and Zaffalon [15,54,55] have developed a general framework for conditioning and updating under incomplete data. It could be applied to the IDM, the derivation of which is based on Bayesian updating of all Dirichlet priors. The approach advocated here in Section 6.1 to derive the predictive probabilities based on coarse data is certainly very similar in spirit and it will be illuminating to explore the exact relationship in detail. Special attention should be paid to a careful study of the representation invariance principle, which is crucial in Walley's argumentation for the IDM. From the considerations at the end of Section 5 it can be concluded that the optimal action may depend on the choice of the sample space, and so the representation invariance principle cannot be extended to the decision theoretic context in a straightforward manner.

Further research should also include a thorough comparison with alternative ways to proceed. This includes firstly other approaches to derive predictive lower and upper probabilities from multinomial observations under lacking prior knowledge [11,52] and secondly an alternative decision theoretic approach where the sampling information is handled by decision functions ([2], see also [29,40,25] for related issues). For the later note that – in the context considered here in contrast to classical subjective expected utility theory based on precise prior probabilities – optimality with respect to posterior expected utility/loss and optimality of decision functions with respect to prior expected risk are no longer necessarily fully compatible.

Acknowledgements

This paper is a revised and extended version of a paper presented at ISIPTA'05 (cf. [16]). We thank two anonymous referees as well as the three referees of the ISIPTA contribution and some participants of the meeting for very valuable comments.

References

[1] T. Augustin, Expected utility within a generalized concept of probability – a comprehensive framework for decision making under ambiguity, Stat. Pap. 43 (1) (2002) 5–22.

[2] T. Augustin, On the suboptimality of robust Bayesian procedures from the decision theoretic point of view – a cautionary note on updating imprecise priors, in: J.M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecises Probabilities and their Applications, Lugano, Switzerland, Carleton Scientific, Waterloo, 2003, pp. 31–45.

[3] T. Augustin, Optimal decisions under complex uncertainty – basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability, Z. Angew. Math. Mech. 84 (10–11) (2004) 678–687.

[4] T. Augustin, Generalized basic probability assignments, Int. J. Gen. Syst. 34 (4) (2005) 451–463.

[5] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, second ed., Springer, New York, 1984.

[6] J.M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecises Probabilities and their Applications, Lugano, Switzerland, Carleton Scientific, Waterloo, 2003.

[7] J.-M. Bernard, An introduction to the imprecise Dirichlet model for multinomial data, Int. J. Approx. Reason. 39 (2–3) (2005) 123–150.
[8] S. Blumenthal, Multinomial sampling with partially categorized data, J. Am. Stat. Assoc. 63 (322) (1968) 542–551.
[9] A. Chateauneuf, M. Cohen, I. Meilijson, New tools to better model behavior under risk and uncertainty: an overview, Finance 18 (4) (1997) 25–46.
[10] A. Chateauneuf, J.Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Moebius inversion, Math. Soc. Sci. 17 (3) (1989) 263–283.
[11] F.P.A. Coolen, T. Augustin, Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, SIPTA, Manno (CH), 2005, pp. 125–135.
[12] G. de Cooman, F.G. Cozman, S. Moral, P. Walley (Eds.), ISIPTA 99, Proceedings of the First International Symposium on Imprecise Probabilities and their Applications, Ghent, Belgium, Imprecise Probability Project, Zwijnaarde (Belgium), 1999.
[13] G. de Cooman, T.L. Fine, T. Seidenfeld (Eds.), ISIPTA'01, Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications, Ithaca, NY, USA, Shaker Publishing, Maastricht, 2001.
[14] G. de Cooman, E. Miranda, I. Couso, Lower previsions induced by multi-valued mappings, Journal of Statistical Planning and Inference 133 (1) (2004) 173–197.
[15] G. de Cooman, M. Zaffalon, Updating beliefs with incomplete observations, Artif. Intell. 159 (1–2) (2004) 75–125.
[16] F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, SIPTA, Manno (CH), 2005.
[17] M. DeGroot, Optimal Statistical Decisions, McGraw-Hill, New York, 1970.
[18] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[19] D. Ellsberg, Risk, ambiguity, and the Savage axioms, Q. J. Econ. 75 (1961) 643–669.
[20] P. Gärdenfors, N.-E. Sahlin, Unreliable probabilities, risk taking, decision making, Synthese 53 (1982) 361–386.
[21] P. Ghirardato, M. Marinacci, Risk, ambiguity, and the separation of utility and beliefs, Math. Oper. Res. 26 (4) (2001) 864–890.
[22] P.H. Giang, P.P. Shenoy, Statistical decisions using likelihood information without prior probabilities, in: Darwiche, N. Friedman (Eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, 2002, pp. 170–178.
[23] P.H. Giang, P.P. Shenoy, Decision making with partially consonant belief functions, in: U. Kjærulff, C. Meek (Eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, 2003, pp. 272–280.
[24] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior, J. Math. Econ. 18 (1989) 141–153.
[25] P.D. Grünwald, J. Halpern, When ignorance is bliss, in: M. Chickering, J. Halpern (Eds.), Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence UAI, Banff, Canada, AUAI Press, Arlington, Virginia, 2004, pp. 226–234.
[26] D.F. Heitjan, D.B. Rubin, Ignorability and coarse data, Ann. Stat. 19 (4) (1991) 2244–2253.
[27] D.F. Heitjan, Ignorability and coarse data: some biomedical examples, Biometrics 49 (4) (1993) 1099–1109.
[28] J.L. Hodges, E. Lehmann, The use of previous experience in reaching statistical decisions, Ann. Math. Stat. 23 (3) (1952) 396–407.
[29] J.Y. Jaffray, Rational decision making with imprecise probabilities, in: G. de Cooman, F.G. Cozman, S. Moral, P. Walley (Eds.), ISIPTA 99, Proceedings of the First International Symposium on Imprecise Probabilities and their Applications, Ghent, Belgium, Imprecise Probability Project, Zwijnaarde (Belgium), 1999, pp. 324–332.
[30] D. Kikuti, F.G. Cozman, C.P. de Campos, Partially ordered preferences in decision trees: computing strategies with imprecision in probabilities, in: R. Brafman, U. Junker (Eds.): Multidisciplinary IJCAI-05 Workshop on Advances in Preference Handling, Edinburgh, Scotland, 2005, pp. 118–123; see also http://wikix.ilog.fr/wiki/pub/Preference05/WsProceedings/Pref05.pdf.

[31] E. Kofler, G. Menges, Entscheidungen bei unvollständiger InformationLecture Notes in Economics and Mathematical Systems, vol. 136, Springer, Berlin, 1976.
[32] E. Kofler, Prognosen und Stabilität bei unvollständiger Information, Campus, Frankfurt/Main, 1989.
[33] V.P. Kuznetsov, Interval Statistical Models, Radio and Communication, Moscow, 1991 (in Russian).
[34] I. Levi, On indeterminate probabilities, Journal of Philosophy 71 (13) (1974) 391–418.
[35] S. Maier, Entscheidung und Analyse bei unvollständiger Information, Diploma thesis, University of Munich, 2004 (in German).
[36] R.F. Nau, Uncertainty aversion with second-order probabilities and utilities, in: G. de Cooman, T.L. Fine, T. Seidenfeld (Eds.), ISIPTA'01, Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications, Ithaca, NY, USA, Shaker, Maastricht, 2001, pp. 273–283.
[37] H.T. Nguyen, E.A. Walker, On decision making using belief functions, in: R.Y. Yager, M. Fedrizzi, J. Kacprzyk (Eds.), Advances in the Dempster–Shafer Theory of Evidence, Wiley, New York, 1994, pp. 311–330.
[38] M. Schervish, T. Seidenfeld, J. Kadane, I. Levi, Extensions of expected utility theory and some limitations of pairwise comparisons, in: J.M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecises Probabilities and their Applications, Lugano, Switzerland, Carleton Scientific, Waterloo, 2003, pp. 496–510.
[39] J. Schubert, On ρ in a decision-theoretic apparatus of Dempster–Shafer theory, Int. J. Approx. Reason. 13 (3) (1995) 185–200.
[40] T. Seidenfeld, A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibility, Synthese 140 (1–2) (2004) 69–88.
[41] T. Seidenfeld, L. Wassermann, Discussion of the paper by Walley (P. Walley, Inferences from multinomial data: learning about a bag of marbles), J. Roy. Stat. Soc. B 58 (1) (1996) 49.
[42] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[43] T.M. Strat, Decision analysis using belief functions, Int. J. Approx. Reason. 4 (5–6) (1990) 391–417.
[44] F. Tonon, A. Bernardini, A. Mammino, Determination of parameter range in rock engineering by means of random set theory, Reliab. Eng. Syst. Safe. 70 (3) (2000) 241–261.
[45] M.C.M. Troffaes, Decision making under uncertainty using imprecise probabilities. Int. J. Approx. Reason., in press, doi:10.1016/j.ijar.2006.06.001.
[46] L.V. Utkin, Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model, Fuzzy Set. Syst. 154 (3) (2005) 413–431.
[47] L.V. Utkin, T. Augustin, Decision making with imprecise second-order probabilities, in: J.M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecises Probabilities and their Applications, Lugano, Switzerland, Carleton Scientific, Waterloo, 2003, pp. 545–559.
[48] L.V. Utkin, T. Augustin, Powerful algorithms for decision making under partial prior information and general ambiguity attitudes, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, SIPTA, Manno (CH), 2005, pp. 349–358.
[49] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[50] P. Walley, Inferences from multinomial data: learning about a bag of marbles (with discussion), J. Roy. Stat. Soc. B 58 (1) (1996) 3–57.
[51] K. Weichselberger, Elementare Grundbegriffe einer allgemeineren Wahrschein-lichkeitsrechnungInterval-lwahrscheinlichkeit als umfassendes Konzept I, Physika, Heidelberg, 2001 (in German).
[52] K. Weichselberger, The logical concept of probability and statistical inference, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, SIPTA, Manno (CH), 2005, pp. 396–405.
[53] R.R. Yager, Decision making under Dempster–Shafer uncertainties, Int. J. Gen. Syst. 20 (3) (1992) 233–245.
[54] M. Zaffalon, Exact credal treatment of missing data, J. Stat. Plan. Infer. 105 (1) (2002) 105–122.
[55] M. Zaffalon, Conservative rules for predictive inference with incomplete data, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, SIPTA, Manno (CH), 2005, pp. 406–415.